Luke MALIK (Osaka University)

Criteria for Identifying Category Mistakes

In this article, I introduce a way to identify the occurrence of a category mistake. Sentences and parts of certain sentences are said to indicate category mistakes. Category mistakes are not identified with sentences and their parts, but with acts. Therefore, I talk about sentences and sentence parts as *indicating* category mistakes. Sentences that are indicative of category mistakes are distinguished from misclassifications, actually false, nomologically false, metaphysically false, and some meaningless sentences. Also, a kind of predicate that cannot figure in category mistakes is identified. Some sentences and sentence parts that are indicative of category mistakes are shown to be meaningless, but others are taken to be merely false. I end by noting some of the ramifications of the criteria presented.

Introduction

Some theorists have found it difficult to provide criteria to distinguish category mistakes from other kinds of error. Prominent contemporary voices avoid providing necessary and sufficient conditions for category mistakes. Whether category mistakes are to be treated syntactically, semantically, or pragmatically is up for debate. Whether category mistakes are meaningless or not is also at issue.[1] Gilbert Ryle, associated most closely with the category mistake, only provides obscure examples, which philosophers are still trying to decipher today (e.g. Goldwater 2018). Still, philosophers and non-philosophers alike freely deploy the concept to criticise their opponents (O'Sullivan 2016.). There is, then, a need to provide criteria that identifies the occurrence of a category mistake and allows the critical application of the concept of that mistake. This I try to do below. The criteria is used to distinguish sentences and sentence parts that are indicative of category mistakes from those that are not. To start with, I explain why I say that sentences and sentence parts are *indicative* of category mistakes.

[1] See Magidor (2013) for an overview.

Sentences or Sentence Parts are *Indicative* of Category Mistakes

Sometimes category mistakes are said to be sentences. For example, this is the first line of the Stanford Encyclopaedia's article on category mistakes:

> Category mistakes *are* sentences such as 'The number two is blue', 'The theory of relativity is eating breakfast', or 'Green ideas sleep furiously' (Magidor 2020, my italics)

This is rejected. First, category mistakes can be associated with subsentential units and criteria will be provided for identifying these.

Second, mistakes are not the objects they result in, but the acts that give rise to those objects (if indeed they give rise to any). We can argue that this is the case with category mistakes by analogy to the spelling mistake. Consider the following typographical error:

(1) Charles Sanders P*ie*rce

It is natural to point to the italicised area in (1) and that this is a spelling mistake. However, this is a metonymic way of talking. The spelling mistake is better identified with the physical act that produced the italicised area rather than the italicised area itself. To see this, consider a writer who types out Peirce's name using a computer keyboard or smartphone. Say, now, (1) appears on the screen. Perhaps, this is due to an underused autocorrect. In any case, it seems that the writer has made *no mistake* despite the incorrect appearance of the name on the computer screen. The writer is not at fault. Suppose, on the other hand, that the writer inputs the name incorrectly (for example, exactly as in (1)). This time we may suppose that the autocorrect compensates for the mistake and the following appears:

(2) Charles Sanders Peirce

The author *has* made a mistake even if there is no *sign* of it. An astute teacher would, certainly, point out the mistake. If we are correct in our assessment, the sign that a mistake is made, for example, the italicised marks in (1), are not entailed nor entail a spelling mistake. The more important point is this, spelling mistakes are not identical to such marks. Rather, based on the examples provided, the spelling mistake is the physical act that can lead to the physical sign of the mistake (or not). The sign can be thought of as an index in Peircian terms (Peirce 1868), a sign of a preceding physical event. Speaking of the sign *as* the mistake is, in

this sense, metonymic. The physical marks that index the mistake are merely *indicative* of a spelling mistake.

In like manner, category mistakes shall not be identified with sentences. Sentences and sentences parts will rather be taken to be *indicative* of a certain kind of mistake: the category mistake. The actual mistake, like the spelling mistake, lies in *putting* bits of language together incorrectly. This is why I shall talk of sentences or sentence parts as *indicative* of category mistakes rather than as category mistakes.

Criteria for Establishing a Sentence or Sentence Part is Indicative of a Category Mistakes: Three Assumptions

A first assumption, here, is that all examples below are to be read *literally*. Many examples below may be read tropically, but this should be avoided.

Second, I assume that we can identify a string of expressions as a sentence or sentence part when, and only when, that string of expressions is combined in the grammatically correct way. A sentence or sentence part, then, is defined grammatically.

In the simplest case to combine a noun with a verb is to form a sentence. For example,

(3) Jasper sleeps

In the next simplest case, a sentence is formed by combining a noun with a verb and the resulting verb phrase with another noun. For example,

(4) Jasper loves biscuits

Generally, both sentences can be associated with a subject predicate form. The first has a one-place predicate, and the second has a two-place predicate.

A third assumption, based on Lowe's (1998) discussion of strict, narrow, and broad logical possibility, is that there are strict, narrow, and broad logical restrictions to think about when considering category mistakes and, in general, how parts of language fit together. So, for example, consider the following sentence:

(5) Jasper is small and hairy

This entails:

(6) Jasper is hairy

This is logically coherent in the strictest sense. *The rules of logic alone guarantee the entailment*.

Consider now (3) again. It seems to entail:

(7) Jasper is not awake

This is logically coherent in the *narrow* logical sense. *The rules of logic, plus the conventional conceptual relations associated with the expressions "asleep" and "awake" guarantee the entailment*.

Lastly, consider:

(8) Water is H2O

This entails:

(9) Necessarily, water is H2O

And:

(10) It is impossible for water to be XYZ[2]

This is logically coherent in the broad logical sense, or in the metaphysical sense. *The rules of logic plus facts about actual ontological relations associated with the actual denotations of "water" and "H2O" guarantee the result*.

These three kinds of restriction will be involved when outlining the criteria for identifying sentences that are indicative of category mistakes.

Before finishing this section, however, I want to focus a little more attention on how these restrictions apply to sentence parts. To understand how we will do this, we must rather refer to the denotations of sentence parts and what they entail. I will assume that the denotations in question are either sets or functions. Combinations of more simple subsentential expressions

[2] Where "XYZ" refers to a very long chemical formula that is not identical to H2O. See Putnam (1975).

that produce more complex subsentential expressions denote sets or functions. The resultant denotations are constituted by the interactions of the denotations of the relevant expressions.

To see how the logical restrictions (strict, narrow, broad) apply to these sentence parts, consider the copula "is." An expression "is" can combine with an adjective, *A*. The former, we shall suppose, denotes a function, **IS**. The latter denotes a set, represented in bold ***A***.[3] These combine semantically like this, with **x** standing for the denotation of an arbitrary entity, and 1 and 0 standing for truth values:

(11) **IS**(**x**, ***A***) = 1 iff **x** is in ***A***, and 0 otherwise

Likewise, for another adjective, *B*:

(12) **IS**(**x**, ***B***) = 1 iff **x** is in ***B***, and 0 otherwise

Suppose, now, it is possible for ***A*** to intersect ***B***, then *A* cannot be a lexical opposite of *B* if *narrow* logical coherence is to be maintained. This is to say, under such constraints, if *A is* a lexical opposite of *B*, the denotation of the former cannot intersect the denotation of the latter.

In case of identifying "is" with an identity function, and an adjective *A* with another function, the denotation of the former applied to the denotation of the adjective, call it g, returns g. And, likewise, if **IS** is applied to the denotation of *B*, call it h, h is returned. If A and B are opposites (like alive/dead), then the denotation of "is" applied to the denotations of *A* and *B*, respectively, *must deliver opposing truth values*. If they do not, *narrow* logical restrictions are violated.

The same is true for the addition of the denotation of "is" with an expression like "H2O." But this time if the denotation **IS** is applied to the denotation of "H2O," it delivers a function, call it **N**, that, when combined with the argument, **water**, *must* deliver 1, and never 0. This is to maintain *broad* logical coherence.

Such relations, strict, narrow, and broad logical relations, temper the understanding of the sentences and sentence parts that we consider and help identify sentences and sentence parts that are indicative of category mistakes. With this in mind, we are now in a position to provide criteria that show when a sentence or part of a sentence is indicative of a category mistake.[4]

---

[3] Below boldface typography will represent the denotation of an expression.

[4] However, footnotes may contain more in depth analysis where needed. These won't be essential to the arguments made in the body of the text but may help with interpreting the relevant formula.

Criteria for Establishing a Sentence or Sentence Part is Indicative of a Category Mistakes

Suppose that the following is a subject-predicate sentence containing a one-place predicate:

(13) SP

Suppose that S stands for a subject expression and P a predicate. Assume next that P belongs to a schema, ϵ, where a schema is a collection of predicates (e.g. P, P', P''...). Each predicate is taken to denote a family of functions (e.g. **P**, **P'**, **P''**...) each of which characterises a set. Each of these sets is a subset of a superset, which we can associate with a predicate P^{+}. An example of a schema is the ordinary colour-schema. Predicates are "red," "green," "blue," etc. Suppose these denote the functions **red**, **green**, **blue**, etc. Applied to entities, they deliver truth values, and characterise sets that fill up the set of things that emit or reflect light.

For convenience sake, let P be a particular predicate of schema ϵ and P^{ϵ} be any other arbitrary predicate of schema ϵ. Then, the criteria for identifying a category mistake for a sentence containing a one-place predicate represented in (13) is given like this:

SP is indicative of a category mistake iff:

(A) $\neg \Diamond$ SP and $\neg \Diamond SP^{\epsilon}$

This is to be justified by the following kind of calculation:

a. SP or SP^{ϵ} entails SP^{+}, but
b. $\neg \Diamond SP^{+}$

That is, either sentence SP or SP^{ϵ} entails SP^{+}, but SP^{+} is logically (strictly, narrowly, or broadly) impossible.

We now state a criteria for establishing when the conjunction of a predicate to an object term indicates a category mistake. The following is to be taken as a sentence containing a two-place predicate, where the sentence is of the simplest kind:

(14) SQO

To say we are speaking of a sentence containing a two-place predicate, where the sentence is

of the simplest kind is to say that the sentence contains: S = subject term denoting an entity; Q = two place predicate; O = object term denoting an entity. Our focus is on the following combination:

(15) QO

Assume now that Q belongs to a schema, ¥, where the schema is a collection of predicates (e.g. Q, Q', Q''...). Again, each predicate is taken to denote a family of functions (e.g. **Q**, **Q'**, **Q"**...) each of which, this time, *delivers a function* that characterises a set. An example of such a schema is a location-schema. Predicates, for example, may be "north of," "east of," "south of," and "west of," denoting functions **north**, **east**, etc. Applied to entities, they deliver functions from entities to truth values. Again, for convenience sake, let Q be a particular predicate of schema ¥ and $Q^{¥}$ be any other arbitrary predicate of schema ¥. Then, the criteria for identifying a category mistake for an expression, QO, containing a two-place predicate is:

QO is indicative of a category mistake iff:

(B) $\mathbf{Q}_{e(et)} + \mathbf{O}_e \neq \mathbf{Q}(\mathbf{O})_{et}$ and $\mathbf{Q}^{¥}_{e(et)} + \mathbf{O}_e \neq \mathbf{Q}^{¥}(\mathbf{O})_{et}$

Justified by the following based on the logical criteria:

a. $\mathbf{Q}_{e(et)} + \mathbf{O}_e = \mathbf{Q}(\mathbf{O})_{et} \rightarrow \perp$
b. $\mathbf{Q}^{¥}_{e(et)} + \mathbf{O}_e = \mathbf{Q}^{¥}(\mathbf{O})_{et} \rightarrow \perp$

Obviously, we need to explain the terminology here. To do this, we need to say a little bit about a certain kind of semantics. This will also allow us to expand on this criteria.

Many semanticists identify expressions with types. Basic types are entities (e) and truth values (t). Expressions, like proper names, denote entities, so are associated with type e. Expressions, like sentences, denote truth values, so are associated with type t. Derived types are built from more basic types and denote functions. A derived type, an et type, denotes a function from entities to truth values. A second derived type, an e(et), type denotes a function from an entity to a function from an entity to a truth value. It is natural to associate a one-place predicate, like an intransitive verb, with an et function; and, a two-place predicate, like a transitive verb, with an e(et) function.

One attraction of this kind of theory is that it allows for "functional application." Consider the following sentence:

(16) Jasper is small

With bold face representing the denotation of each expression and subscripts representing the type of each denotation, we can present (16) like so:

(17) $\mathbf{Jasper}_{e} + \mathbf{IS}_{et(et)} + \mathbf{small}_{et}$

Here the subscripts tell us the *type* of the denotation in question *and* tell us, also, how each denotation interacts with the other denotations. The capitalisation signifies that the function in question is not arbitrary but must be defined in a specific way.

Yoad Winter (2016) defines the kind of interaction in question like this:

Applying a function F of type τσ to an object of x of type $\boldsymbol{\tau}$ *gives an object F (x) of type σ In short:*

Types: $(\tau\sigma) + \tau = \tau + (\tau\sigma) = \sigma$
Denotations: $F_{\tau\sigma} + x_{\tau} = x_{\tau} + F_{\tau\sigma} = F(x)$: σ

Functional application on (18), then, looks then like this:

(18) $\mathbf{Jasper}_{e} + \mathbf{IS}_{et(et)} + \mathbf{small}_{et} = \mathbf{Jasper}_{e} + \mathbf{small}_{et} = \mathbf{small}(\mathbf{Jasper})$: t

The function associated with **IS** accepts an et function and delivers an et function. In (18), the function accepts the function associated with the denotation of "small" *and* delivers that very same function. The function **small** takes an entity and delivers a truth value. In the example, the entity associated with the denotation **Jasper**.

Heim and Kratzer (1998), noting Davidson's influence, model meaning in a way that exemplifies the relevant truth conditions (1, 0). Amalgamating this kind of modelling with the representation system above, and working things vertically, we get something like this:

1. $\mathbf{Jasper}_{e} + \mathbf{is}_{et(et)} + \mathbf{small}_{et}$
2. $= \mathbf{Jasper}_{e} + \mathbf{is}_{et(et)} + \mathbf{small}_{et}$
3. $= \mathbf{Jasper}_{e} + \mathbf{small}_{et}$
4. = 1 iff Jasper is small

The denotations of functions are made more transparent by using the lambda notation

convention, which gives us this:

1. $\mathbf{Jasper}_{e} + \mathbf{is}_{et(et)} + \mathbf{small}_{er}$
2. $= \mathbf{Jasper}_{e} + \lambda g_{et} \,.\, g_{et} + \lambda x_{e} \,.$ x is small
3. $= \mathbf{Jasper}_{e} + \lambda x_{e} \,.$ x is small
4. = 1 iff Jasper is small

"$\lambda g_{et} \,.\, g_{et}$" is read "the unique et(et) function that maps an et function to itself." "$\lambda x_{e} \,.$ x is small" is read "the unique et function that maps an entity to the truth value 1 iff x is small, and 0 otherwise."

I would like to appropriate this descriptive notation in order to describe the kind of category mistake in question. Consider, again, the following representation of a sentence containing a two-place predicate, which is of the simplest kind:

(19) SQO

Appropriating the conventions and notations introduced, we may analyse (19) like this:

1. $\mathbf{S}_{e} + \mathbf{Q}_{e(et)} + \mathbf{O}_{e}$
2. $= \mathbf{S}_{e} + [\lambda x_{e} \,.\, \lambda y_{e} \,.$ y Q x$] + \mathbf{O}_{e}$
3. $= \mathbf{S}_{e} + \lambda y_{e} \,.$ y Q O
4. = 1 iff SQO

If the functional application breaks down with the move from step 2 to step 3 because there is no function described $\lambda y_{e} \,.$ x Q O, and there is no alternative function described $\lambda y_{e} \,.$ x Q¥ O, then putting Q together with O is indicative of a category mistake. The combination is, further, *meaningless* because *it denotes nothing*. And it renders a sentence in which it is contained meaningless because it denotes nothing.

But what does it mean to say there is no function of the kind in question? Logical restrictions, broad and narrow, will restrict our choices here. So, for example, to put **Q** together with **O** entails that the denotation of **O** has certain properties. Given our conceptual and metaphysical knowledge about **O**, this is absurd. If we want to reject absurd conclusions, then the denotation of Q cannot be applied to the denotation of O. This justifies thinking the denotation of Q combined with the denotation of O results in no further function. This is expressed like this:

QO is indicative of a category mistake iff:

(B) $\mathbf{Q}_{e(et)} + \mathbf{O}_e \neq \mathbf{Q}(\mathbf{O})_{et}$ and $\mathbf{Q}^{¥}_{e(et)} + \mathbf{O}_e \neq \mathbf{Q}^{¥}(\mathbf{O})_{et}$

Justified by the following based on our logical restrictions:

a. $\mathbf{Q}_{e(et)} + \mathbf{O}_e = \mathbf{Q}(\mathbf{O})_{et} \rightarrow \perp$
b. $\mathbf{Q}^{¥}_{e(et)} + \mathbf{O}_e = \mathbf{Q}^{¥}(\mathbf{O})_{et} \rightarrow \perp$

That is, the functional application of the two-place predicate to the entity in question is absurd. Examples are provided below.

In opposition to the criteria that identify sentences or sentence parts as indicative of category mistake, we can say when sentences and sentence parts are *not* indicative of category mistakes. For sentences with one-place predicates:

(C) $\Diamond$ SP or $\Diamond$ SP$^{€}$

For a sentence containing a two-place predicate, which is of the simplest kind:

(D) $\mathbf{Q}_{e(et)} + \mathbf{O}_e = \mathbf{Q}(\mathbf{O})_{et}$ or $\mathbf{Q}^{¥}_{e(et)} + \mathbf{O}_e = \mathbf{Q}^{¥}(\mathbf{O})_{et}$

We can now generalise things, which will be important to us below. Consider a grammatically possible combination of expressions:

(20) XY

We will suppose that X stands for some expression that denotes a function **X** and belongs to a schema, K. And that X' stands for some arbitrary alternative expression that denotes some alternative function **X'**, and which also belongs to the schema, K. EIther of **X** or **X'** delivers a function. An example of this might be a schema that contains a two-place predicate, a three-place predicate, a second-order predicate, a modifying predicate, etc. Given this, the relevant sentence parts are indicative of category mistakes iff:

(E) $\mathbf{X}_{\tau\sigma} + \mathbf{Y}_{\tau} \neq \mathbf{X}(\mathbf{Y})_{\sigma}$ and $\mathbf{X'}_{\tau\sigma} + \mathbf{Y}_{\tau} \neq \mathbf{X'}(\mathbf{Y})_{\sigma}$

Justified by the following based on the logical criteria :

a. $\mathbf{X}_{\tau\sigma} + \mathbf{Y}_{\tau} \neq \mathbf{X}(\mathbf{Y})_{t} \rightarrow \perp$
b. $\mathbf{X'}_{\tau\sigma} + \mathbf{Y}_{\tau} \neq \mathbf{X'}(\mathbf{Y})_{\sigma} \rightarrow \perp$

And, sentence parts that are not indicative of category mistakes are, then, described like this:

(F) $\mathbf{X}_{\tau\sigma} + \mathbf{Y}_{\tau} = \mathbf{X}(\mathbf{Y})_{\sigma}$ or $\mathbf{X'}_{\tau\sigma} + \mathbf{Y}_{\tau} = \mathbf{X'}(\mathbf{Y})_{\sigma}$

This is to say, when functional application collapses, based on extra-semantic logical entailments, we have reason to think that a combination of terms is indicative of a category mistake.

We use (A) - (F) to identify sentences or parts of sentences indicative of category mistakes below. This can be viewed as a test for our criteria.

Sentences and Sentence Parts That Are Not Indicative of Category Mistakes

Consider the following sentences:

(21) Whales are fish
(22) Sherlock Holmes is dead
(23) Professor Moriarty killed Sherlock Holmes
(24) Circles are square
(25) Water is NaCl
(26) He is pregnant
(27) A square circle is beautiful
(28) I drew a square circle
(29) Category mistakes are interesting

These sentences *are not* indicative of category mistakes by the criteria above.

Consider (21). The predicate belongs to the following schema: the biological schema of predicates that apply to vertebrates. This includes the predicates “amphibian,” “bird,” “fish,” “mammal,” “reptile.” Though (21) is false, the following is true:

(30) Whales are mammals

Thus, we can say, according to (A) or (C), (21) is not a category mistake. Rather, (21) is a

mere misclassification.

Consider, now, (22). This is not a category mistake because it is a sentence that is possibly true. Broadly speaking, (22) may be true and, thus, according to (A) is not a category mistake. Although, it is *actually* false

Consider (23). This sentence, too, may be true. To say that, however, assumes that the two-place predicate denotes a function when combined with the object term and that that is not ruled out logically. It isn't.

The semantics that would make sense of this is an expanded version of the semantics presented above. Expressions like "Sherlock Holmes" are proper nouns. But expressions like "Sherlock Holmes" denote no entity in the actual world, though it doesn't seem logically incoherent to think they denote entities in possible worlds. In a functional semantics such expressions are said to denote "individual concepts." The type an individual concept gets in such a semantics is (s,e), which is a function from possible worlds (s) to entities (e). A one-place predicate gets the type (se(t)), and a two-place expression (like "kill") denotes a function of the type (se (se(t))). There is nothing *logically* incoherent, in the broad sense, with supposing that the denotation of "kill" can combine with the denotation of "Sherlock Holmes" resulting in a denotation for a one-place predicate "kill Sherlock Holmes." Thus, even though (D) does not allow us to rule out a category mistake for the relevant combination of parts in (3), (F) will (with the types appropriately rendered).

Turn now to (24), assuming that the predicate belongs to a mathematical schema that classifies shapes, we can provide a sentence that is true by drawing on another predicate of that schema:

(31) Circles are circles

Thus, though, (24) is logically impossible, it is not indicative of a category mistake by the criteria set out above.

How about (25)? Assuming that "NaCl" belongs to a schema of chemical composition predicates, we can, again, provide a sentence that is true drawing on another predicate of that schema:

(32) Water is H20

Thus, though, (25) is metaphysically impossible, it is not indicative of a category mistake by the criteria set out above.

Next, assuming that it is not logically impossible, even if it is biologically impossible,

that men can be pregnant, (26) is not a category mistake. So, (even) if (26) is *nomologically* impossible, it is not indicative of a category mistake.

Next, (27) and (28) do not appear to be category mistakes at the sentential levelbecause in (27) the subject expression does not denote, and in (28) the predicate expression does not denote. In this sense, both sentences are just meaningless.

However, we might wonder whether the subject term in (27), and thus the predicate expression in (28), are indicators of category mistakes. That is, is "square circle" indicative of a category mistake? We can suppose that either (E) or (F) are relevant to this question. Suppose the denotations of "square" here is $\mathbf{square}^{\mathbf{mod}}{}_{\tau\sigma}$ and "circle" is $\mathbf{circle}_{\sigma}$. In effect, the former would be an identity function, like **IS**, that returns the entity or function that it is applied to. Further, we can suppose that "square" is a modifier, and it belongs to a modifier-schema to which other expressions of the following kind belong "circular," "elliptic," "triangular," etc. Each denotes a function, $\mathbf{circular}^{\mathbf{mod}}{}_{\tau\sigma}$ $\mathbf{ellipitic}^{\mathbf{mod}}{}_{\tau\sigma}$ $\mathbf{triangular}^{\mathbf{mod}}{}_{\tau\sigma}$ etc. Then, since, for example, $\mathbf{circular}^{\mathbf{mod}}{}_{\tau\sigma}$ can combine with $\mathbf{circle}_{\sigma}$ and return an argument for another expression, "square circle" is not to be thought of as indicative of a category mistake. Perhaps, that is clearer given the following example:

(33) A perfectly circular circle is beautiful

Although, for sure, a "square circle" denotes nothing since square circles are logically impossible, by our lights, it is not indicative of a category mistake.

Last, consider (29). The predicate expresses a subjective impression. Other expressions that express subjective impressions are often "ing" ending adjectives, like "disgusting," "worrying," etc. We can, thus, associate these adjectives with predicates that belong to what we might call the "subjective impressions schema " (SIS). SIS expressions don't seem to figure in category mistakes. These predicates combining with subject terms denoting entities (or whatever) will always produce sentences that are possibly true. Thus, (29) is not a category mistake. And, assuming grammatical compliance, if used in a modifying manner, it will always be possible to have them denote a function that returns its argument. The real challenge is to allow them to shift type in order to accommodate the expressions they are paired with, which seem to be unrestricted.

To sum up, misclassifications, actually false, logically false, metaphysically false, and nomologically false sentences, are not automatically sentences that are indicative of category mistakes. Neither are sentences with non-referring parts. Last, some predicates and predicate schemas cannot be used in a way that produces a sentence or sentence part that is indicative of a category mistake.

Sentences and Sentence Parts That Are Indicative of Category Mistakes

Consider next though:

(34) Nixon is a square
(35) Biden smoked Trump
(36) The container is inside itself

One sentence *is* indicative of category mistakes; the others contain a part which is indicative of the mistake.

Take (34), we may assume that the predicate belongs to a the shape-predicate schema, all predicates of which are defined mathematically. Relative to the predicates of this schema, (34) is false and any alternative to (34) produced by replacing the offending predicate is false. We, thus, conclude by (A), that (34) is a category mistake.

Let's now think about (35). We can think about the sentence as above:

1. $\mathbf{Biden}_e$ + $\mathbf{smoked}_{e(et)}$ + $\mathbf{Trump}_e$
2. = $\mathbf{Biden}_e$ + $[\lambda x_e . \lambda y_e .$y smoked x$]$ + $\mathbf{Trump}_e$
3. = $\mathbf{Biden}_e$ + $\lambda y_e .$y smoked Trump
4. = 1 iff Biden smoked Trump

The step from 2 to 3 is illegitimate, for it entails Trump can be inhaled and exhaled, and this is absurd. Now, supposing that "smoke" is a predicate of a schema of similar predicates, for example, a schema including predicates like "breathe in," "inhale," "toke," "vape," etc., there is no alternative that can be swapped in for "smoke" without leading to absurdity. According to our criteria, then, this means "smoked Trump" is indicative of a category mistake. Moreover, since the denotation of the expression "smoked" cannot be combined with the denotation of "Trump," in order to avoid absurdity, the expression "smoked Trumped" denotes nothing. And, thus, (35) is meaningless.

Next, consider (36). For this sentence, we suppose that "inside" is associated with a predicate from a two-place predicate binary predicate schema. The other predicate of this schema is "outside." The first predicate denotes an e(et) function **inside**, and the other predicate of the schema denotes an e(et) function **outside**. The reflexive pronoun is defined by giving it the following lambda expression $\lambda R_{e(et)} . \lambda x_e .$ x is R x. With this mind, and supposing that "the" is vacuous, we can analyse (36) like this:

1. **The container**$_{e}$ + **IS**$_{et(et)}$ + **inside**$_{e(et)}$ + **ITSELF**$_{(e(et))(et)}$
2. = **The container**$_{e}$ + λg_{et} . g + **inside**$_{e(et)}$ + $\lambda R_{e(et)}$. λx_{e}. x is R x
3. = **The container**$_{e}$ + λg_{et} . g + λx_{e} . x is inside x
4. = **The container**$_{e}$ + λx_{e} . x is inside x
5. = 1 iff the container is inside the container

The analysis breaks down when we move from step 2 to 3 because it is logically impossible for one thing to be inside itself. The alternative to (36):

(37) The container is outside itself

will, when analysed semantically, break down at exactly the same point. We can conclude that the combination of "inside" and "itself" does not produce a legitimate value and, therefore, we draw two conclusions: (a) "inside itself" is indicative of a category mistake, and (b) (36) is meaningless.

Conclusion

I have provided criteria that establish when a sentence or sentence part is indicative of a category mistake, and help to distinguish sentences and parts that are indicative of category mistakes and those that are not. We have applied the criteria to some sample sentences and parts and, hopefully, the way sentences have been sorted by the criteria has met intuitions about what sentences are and are not category mistakes. The next step is to apply the criteria to some more famous examples and draw connections with other theories of category mistakes (e.g. Pap's (1960) account of category mistakes, whose notion of "predicate families" seem to be close to the schemas mentioned above).

For the moment, we can note a couple of things. First, type mismatches and category mistakes do not seem to be aligned. A type mismatch is of the following kind:

(38) e + et(et) + t

This is a different kind of mistake to the one we are defining because in our analysis the types were not mismatched. Also, consider the following two sentences:

(39) 2 is interesting

(40) Jasper is interesting

Perhaps, these are analysed like this:

(41) $\mathbf{2}_{n}$ + $\mathbf{is}_{et(et)}$ + $\mathbf{interesting}_{et}$
(42) $\mathbf{Jasper}_{e}$ + $\mathbf{is}_{et(et)}$ + $\mathbf{interesting}_{et}$

(41) seems to be type mismatched (where n is a number), but (42) doesn't. On our criteria for associating a category mistake with subject predicate sentence, both (39) and (40) are *not* category mistakes. The following sentences may present more of a problem:

(43) 2 is an interesting prime
(44) Jasper is an interesting cat

That is, since the modifying expression here is paired with expressions that have denotations of different types. For example, (where the type nt is a function from numbers to truth values):

(45) ... + $\mathbf{interesting}^{\mathbf{mod}}_{et(et)}$ + $\mathbf{prime}_{nt}$
(46) ... + $\mathbf{interesting}^{\mathbf{mod}}_{et(et)}$ + $\mathbf{cat}_{et}$

Perhaps, then, "interesting" should be taken to denote different functions according to the expression it is paired with. Still, we have here examples of *type-mismatches*, not category mistakes according to our criteria.

A last point, it may be argued that the criteria we have presented do not actually characterise category mistakes and that type mismatches are better associated with characterising category mistakes (after all what Ryle (1938) called type mismatches did latter seem to become Ryle's (1949) category mistakes). One response to this is that the criteria do group the sentences and sentence parts that we do want to think of as indicative of category mistakes. Another response is to accept the complaint and take the criteria as grouping a new kind of mistake, perhaps, we can call it a schema-mistake.

Bibliography

Goldwater, J. B. P. 2018. "Ryle, the Double Counting Problem, and the Logical Form of a

Category Mistake." *The Journal of the History of Philosophy*, 56 (2). 337-359.

Heim, I. and Kratzer, A. 1998. *Semantics in Generative Grammar*. Blackwell.

Lowe, E. J. 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. Clarendon Press.

Magidor, O. 2013. *Category Mistakes*. Oxford University Press.

Magidor, O. 2020. "Category Mistakes", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.),
forthcoming URL = <https://plato.stanford.edu/archives/win2020/entries/category-mistakes/>

O'Sullivan, L. "The Idea of a Category Mistake: From Ryle to Habermas, and Beyond. *History of European Ideas*, 42:2. 178-194.

Pap, A. 1960. "Types and Meaninglessness." *Mind 69* (273): 41-54???

Peirce, C. S. 1868. "On a New List of Categories." *Proceedings of the American Academy of Arts and Sciences*. 287-298.

Putnam, H. 1975. "The Meaning of "Meaning."" *Language, Mind, and Knowledge, Minnesota Studies in the Philosophy of Science,* Volume 7: 131-193.

Ryle, G. 1938. "Categories." *Proceedings of the Aristotelian Society*, Vol. XXXVIII.

Ryle, G. 1949. *The Concept of the Mind*. Hutchinson.

Winter, Y. 2016. *Elements of Formal Semantics: An Introduction to the Mathematical Theory of Meaning in Natural Language.* Edinburgh University Press.